

REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-04-

0384

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing existing information, gathering material and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE 19 November 03	3. REPORT TYPE AND DATES COVERED Final Report 12/1/98 through 11/30/01
4. TITLE AND SUBTITLE Compression and Networking Issues Related to Robust Transmission of Video over Packet Switched Networks		5. FUNDING NUMBERS #F49620-99-1-0062	
6. AUTHOR(S) Professor Avideh Zakhor			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California, Berkeley Electronics Research Laboratory 253 Cory Hall Berkeley, CA 94720		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Air Force Office of Scientific Research 801 N. Randolph St Room 732 Arlington, VA 22203-1977		10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.			
12 a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		12 b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) With ever growing network resources, video communication is an important part of today's internet applications. In this research, we developed a spectrum of techniques for unicast and multicast video communication over packet switched networks. These included best effort networks, as well as Quality of Service (QoS) enabled ones, such as those with Diffserv. In doing so, we pay particular attention to video compression techniques, as well as networking issues. Specifically, in the area of video compression, we continued our efforts on application of overcomplete signal expansion schemes to video coding, and explored a number of issues related to matching pursuits based video coding. These include, dictionary approximation, dictionary design, modulus quantization, rate control, multiple description coding and its application to wireless video communication. In the area of video multicast, we proposed the use of hierarchical FEC as an error control mechanism that allows receivers to individually trade off latency for received video quality. This scheme is efficient since FEC packets are used to protect only the more important data layers and are multicast only to receivers that need them, thereby improving network utilization. We perform actual MBONE experiments to evaluate the performance of our scheme.			
14. SUBJECT TERMS		15. NUMBER OF PAGES	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

20040804 062

mm

12/18/03

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)**FINAL REPORT****Principal Investigators:** Professor Avidah Zakhor

231 Cory Hall

Department of Electrical Engineering

University of California

Berkeley, CA 94720

Grant Number: F49620-99-1-0062**Reporting Period:** 12/1/98 through 11/30/01**Title:** Compression and Networking Issues related to Robust
Transmission of Video over Packet Switched Networks**1 Introduction**

Image and video compression algorithms are an important part of many transmissions and storage systems. The main goal of this progress report is to summarize our findings over the past three years in the area of video compression and transmission over packet switched networks.

2 Video Communication over Packet Switched Networks

With ever growing network resources, video communication is already an important part of today's Internet applications. Over the past five years, we have developed novel techniques for unicast and multicast video transmission over today's best effort networks. Our approach has been to address compression and networking issues jointly, rather than two separate, disjoint problems. In what follows, I describe our findings in the area of video communication over packet switched networks [2, 8, 14, 20, 27].

**2.1 Packet Classification Schemes for Streaming Video over
Delay and Loss Differentiated Networks**

Differentiated services (DiffServ) has been under investigation by IETF to provide relatively simple and coarse traffic differentiation in the Internet. In the past few years, we have developed a framework for transmitting MPEG video by dividing the bit-stream into sub-streams of different delay and loss requirements [14]. The sub-streams are then transported using multiple DiffServ traffic classes of different bandwidth, transmission delay, and packet loss characteristics. The use of multiple traffic classes to carry video improves network utilization as packets are transmitted in traffic classes with QoS commensurate to their requirements, rather than all in the "best" class. This is related to the existing use of scalable coding for loss-differentiated networks. However, in this work, we consider

non-scalable MPEG video due to the abundance of existing content, and also consider loss and delay differentiations simultaneously. We have developed a number of packet classification schemes for MPEG bit-stream based on delay and loss characteristics of the data, and compared them using commercial DVD content. We have simulated transmission of DVD movies over a DiffServ enabled network, and shown a distortion reduction of over 4 dB using a packet classification scheme optimized for loss. We have developed another packet classification scheme optimized for delay which reduces end-to-end playback time by 30 ms as compared to packet classifiers that treat the MPEG stream as homogeneous.

2.2 Error Control for Video Multicast using Hierarchical FEC

Bit-rate scalable video compression with layered multicast has been shown to be an effective method to achieve rate control in heterogeneous networks. Over the past few years, we have advocated the use of hierarchical FEC as an error control mechanism that allows receivers to individually trade-off latency for received video quality [2, 20, 27]. The scheme is efficient since FEC packets are used to protect only the more important data layers and is multicast only to receivers that need them, thereby improving network utilization. Furthermore, there is no loss in error correcting capability by using hierarchical FEC when maximum distance separable codes are used. Actual MBONE experiments were performed to evaluate the performance of the proposed scheme.

3 Signal Decomposition on Over-complete Basis with Applications to Video Coding

Video compression is important in many applications, including video telephony, streaming video over the internet and wireless video communication systems where bandwidth is a premium. Since 1992, we have focused our efforts on developing a new class of low bit rate video compression algorithms based on over-complete signal expansion techniques such as matching pursuit (MP). Over-complete signal decomposition using matching pursuits has been shown to be an efficient technique for coding motion residual images in a hybrid video coder. In what follows, I will outline highlights of our efforts on matching pursuits based video coding over the past three years [4, 9, 16, 23, 26, 11, 12].

3.1 Dictionary Approximation for Matching Pursuit Video Coding

Dictionary design is an important issue for matching pursuits based video coding system, and others have shown alternate dictionaries which lead to either coding efficiency improvements or reduced encoder complexity. Over the past few years, we have introduced for the first time a design methodology which incorporates both coding efficiency and complexity in a systematic way[16]. The key to our new method is an algorithm which takes an arbitrary 2-D dictionary and generates approximations of the dictionary which have fast 2-stage implementations. By varying the quality of the approximation, we can explore a systematic tradeoff between the coding efficiency and complexity of the matching pursuit video encoder. As a practical result, we show cases where complexity is reduced by a factor of 500 to 1000 in exchange for small coding efficiency losses of around 0.1 dB PSNR.

3.2 Modulus Quantization for Matching Pursuit Video Coding

Unlike orthogonal decomposition, matching pursuit uses an in-the-loop modulus quantizer which must be specified before coding begins. This complicates the quantizer design, since the optimal quantizer depends on the statistics of the matching pursuit coefficients which in turn depend on the in-loop quantizer actually used. Over the past few years, we have addressed the modulus quantizer design issue, specifically developing frame-adaptive

quantization schemes for the matching pursuit video coder [4, 23, 26]. Adaptive dead-zone subtraction is shown to reduce the information content of the modulus source, and a uniform threshold quantizer is shown to be optimal for the resulting source. Practical 2-pass and 1-pass algorithms are developed to jointly determine the quantizer parameters and the number of coded basis functions in order to minimize coding distortion for a given rate. The compromise 1-pass scheme performs nearly as well as the full 2-pass algorithm, but with the same complexity as a fixed quantizer design. The adaptive schemes are shown to outperform the fixed quantizer used in earlier works, especially at high bit rates where the gain is as high as 1.7 dB.

3.3 Matching Pursuits Multiple Description Coding for Wireless Video

Multiple description coding (MDC) is an error resilient source coding scheme that creates multiple bit-streams of approximately equal importance. Over the past few years, we have developed a 2 description video coding scheme based on the 3 loop structure proposed earlier [11]. We modify the discrete cosine transform structure to the matching pursuits framework and evaluate performance gain using maximum likelihood (ML) enhancement when both descriptions are available. We find that ML enhancement works best for low motion sequences. Performance comparison is made between our MDC scheme and single description coding (SDC) schemes over two-state Markov channels and Rayleigh fading channels. We find that MDC outperforms SDC in bursty slowly varying environments. In the case of Rayleigh fading channels, interleaving helps SDC close the gap and even outperform MDC depending on the amount of interleaving performed, at the expense of additional delay.

3.4 Learning Dictionaries for Matching Pursuits Based Video Coders

Over the past few years, we have developed a learning scheme for designing dictionaries of two-dimensional functions for matching pursuits (MP) based video coding [12]. The motivation is to improve the performance of such codecs by adapting the structure of the dictionary functions to specific bit-rates of types of sequences. The scheme we have developed is based on vector quantization (VQ), and uses an inner-product based distortion measure. The different processing steps, consist of data extraction from the motion compensated error frames, training, pruning, and testing. We have found that for high bit-rate QCIF sequences we can achieve improvements of up to 0.66 dB.

3.5 Rate Control Layered Video Compression Using Matching Pursuits

Over the past few years, we have developed a multi-pass rate control scheme for SNR scalable encoding based on MP [21]. The rate control algorithm enforces constant quality on each layer, while keeping the bit budget for each layer at a pre-specified target level. We formulate this as a zero finding problem, and solve it using Newton's method. Experimental results on 14 video sequences are included, showing that layered video can be encoded at constant quality in about 5 encoding iterations per layer, while satisfying bit budget constraints with 1:5/tolerance.

4 Content Analysis of Web Multimedia Documents

The amount of information on the World Wide Web has grown enormously since its creation in 1990. Since there is no central management on the web, duplication of content is inevitable. As reported by Shivakumar and Garcia-Molina in 1998, about 46% of all the text documents on the web have at least one "near-duplicate" { document which is identical except for low level details such as formatting. The problem is more severe for videos as they are often mirrored in multiple locations, formats and bitrates to facilitate

downloading and streaming. Multimedia authoring tools also enable users to slightly modify existing video clips and to republish them on the web. An efficient algorithm to identify similar videos can therefore be beneficial to many web retrieval scenarios such as presenting uncluttered search results, and providing alternatives in the case of expired links or network outages.

Over the past few years, we have developed an efficient algorithm called video signature to detect similar video sequences for large databases such as the web. The idea is to first form a "signature" for each video sequence by selecting a small number of its frames that are most similar to a number of randomly chosen seed images[15, 19]. Then the similarity between any two video sequences can be reliability estimated by comparing their respective signatures. Using this method, we achieve 85% recall and precision ratios on a test data based of 377 video sequences. As a proof of concept, we have applied our proposed algorithm to a collection of 1800 hours of video corresponding to around 45000 clips from the web. Our results indicate that, on average, every video in our collection from the web has around five similar objects.

We have also developed a new signature clustering algorithm to further improve retrieval performance[10]. The algorithm treats all the signatures as an abstract threshold graph, where the threshold is determined based on local data statistics. Similar clusters are identified as highly connected regions in the graph. This algorithm outperforms simple thresholding and hierarchical clustering techniques in identifying a set of manually-determined similar clusters from a dataset of 46,356 web video clips. At 95% precision, our algorithm attains 85% while simple thresholding and complete-link hierarchical scheme attain 67% to the entire dataset, 6,900 similar clusters are identified, with an average cluster size of 2.81 video clips. The distribution of cluster sizes follows a power-law distribution, which has been shown to describe many web phenomena.

5 Interactions/Transitions

The Matching Pursuits research described in section 3 has been patented at U.C. Berkeley and been available for licensing to companies since 1998. Truvideo has licensed this technology from U.C. Berkeley and is currently commercializing it. As of 2003, matching pursuits based video is live on Verizon Wireless, U.S. Cell, and Alltel in the U.S. and on Hutchison in Thailand.

6 Publications

Below is the list of publications resulting from this grant. Specifically this grant resulted in 9 journal publications, and 18 conference publications.

References

- [1] Nelson L. Chang and Avidesh Zakhor, "Constructing a Multi-valued Representation for View Synthesis," International Journal of Computer Vision, 45(2):157-190, November 2001.
- [2] W. Tan and A. Zakhor, "Video Multicast using Layered FEC and Scalable Compression" IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, March 2001, pp. 373- 386.

- [3] N. Dimitrova, H. Yu, F. Galliano, R. Koenen, A. Zakhor, C. Bouman, "Entry into the Content Forest: The Role of Multimedia Portals", IEEE Multimedia, Vol. 7, No. 3, July-Sep 2000, pp. 14 - 20.
- [4] R. Neff and A. Zakhor, "Modulus Quantization for Matching Pursuit Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 6, September 2000, pp. 895-912.
- [5] G. Cheung and A. Zakhor, "Optimal Bit Allocation Strategy for Joint Source/Channel Coding of Scalable Video", IEEE Transactions on Image Processing, Vol. 9, No. 3, pp. 340 - 357, March 2000.
- [6] Emile Sahouria and A. Zakhor, "Content Analysis of Video Using Principal Components" IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 8, pp. 1290-1298, December 1999.
- [7] N. S. Shah and A. Zakhor, "Resolution Enhancement of Color Video Sequences," IEEE Transactions on Image Processing, June 1999, vol. 8, no. 6, pp. 879-885.
- [8] W. Tan and A. Zakhor, "Real-Time Internet Video Using Error Resilient Scalable Compression and TCP-Friendly Transport Protocol," IEEE Transactions on Multimedia, June 1999, vol. 1, no. 2, pp. 172-186.
- [9] O.K. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor, "Video Compression Using Matching Pursuits," IEEE Transactions on Circuits and Systems for Video Technology, February 1999, vol. 9, no. 1, pp. 123-143.

Conference Papers

- [10] S. Cheung and A. Zakhor, "Video similarity detection with video signature clustering" in Proceedings of the International Conference on Image Processing, Thessaloniki, Greece, October 2001, Vol. 2, pp. 649 652.
- [11] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video" in Proceedings of the International Conference on Image Processing, Thessaloniki, Greece, October 2001, Vol. 1, pp. 926 - 929.
- [12] P. Schmid, and A. Zakhor, "Learning dictionaries for matching pursuits based video coders", in Proceedings of International Conference on Image Processing, Thessaloniki, Greece, October 2001, Vol. 3, pp. 562 - 565.
- [13] C. Frueh and A. Zakhor, "Fast 3D model generation in urban environments", in International Conference on Multisensor Fusion and Integration for Intelligent Systems 2001, Baden-Baden, Germany, August 2001, pp. 165-170.
- [14] W. Tan and A. Zakhor, "Packet classification schemes for streaming MPEG video over delay and loss differentiated networks" in Proceedings of Packet VideoWorkshop 2001, Kyongju, Korea, April 2001.
- [15] S. Cheung and A. Zakhor, "Efficient video similarity measurement and search," International Conference on Image Processing, Vancouver, Canada. September 10-13, 2000, Vol. I, pp. 85-89.

[16] R. Neff and A. Zakhor, "Dictionary Approximation for Matching Pursuit Video Coding" in Proceedings of the International Conference on Image Processing, Vancouver, Canada, September 2000, Vol. 2, pp. 828-831.

[17] T. Nguyen and A. Zakhor, "Performance Analysis of an H.263 Video Encoder for IRAM", Proceedings of International Conference on Image Processing, Sep. 2000, Vancouver, Canada, Vol. 3, pp. 98- 101.

[18] Vito Dai and Avidesh Zakhor, "Lossless Layout Compression for Maskless Lithography Systems" Emerging Lithographic Technologies IV, Proceedings of the SPIE Vol. 3997, pp. 467-477, San Jose, CA, March, 2000.

[19] S. Cheung and A. Zakhor. "Estimation of Web Video Multiplicity". Proceedings of SPIE, Internet Imaging, SPIE volume 3964, pp. 34-46. San Jose, California. Jan 22-28, 2000.

[20] Wai-Tian Tan and A. Zakhor, "Error Control for Video Multicast using Hierarchical FEC", Proceedings of the International Conference on Image Processing, Kobe, Japan, October 1999, Vol. 1, pp. 401-405.

[21] E. Miloslavsky and A. Zakhor, "Rate Control for Layered Video Compression Using Matching Pursuits," in Proceedings of the International Conference on Image Processing, Kobe, Japan, October 1999, Vol. 2, pp. 357 - 361

[22] Nelson Chang and Avidesh Zakhor, "A Multi-valued Representation for View Synthesis," in Proceedings of the International Conference on Image Processing, Kobe, Japan, October 1999, Vol. 2, pp. 505-509

[23] Ralph Neff and Avidesh Zakhor, "Adaptive Modulus Quantizer Design for Matching Pursuit Video Coding," in Proceedings of the International Conference on Image Processing, Kobe, Japan, October 1999, Vol. 2, pp. 81-85

[24] E. Sahouria and A. Zakhor, "A Trajectory Based Video Indexing System For Street Surveillance," in Proceedings of the IEEE International Conference on Image Processing (ICIP), 1999. [Postscript]

[25] N. L. Chang and A. Zakhor, "A Multivalued Representation for View Synthesis," in Proceedings of the International Conference on Image Processing, Kobe, Japan, October 1999, vol. 2, pp. 505-509.

[26] R. Neff and A. Zakhor, "Adaptive Modulus Quantizer Design for Matching Pursuit Video Coding," in Proceedings of the International Conference on Image Processing, Kobe, Japan, October 1999, vol. 2, pp. 81-85.

[27] W. Tan and A. Zakhor, "Multicast Transmission of Scalable Video using Receiver-driven Hierarchical FEC ," in Packet Video Workshop, New York City, New York, April 1999.